Intra-site spatial analysis in ethnoarchaeology

Carla Lancelotti[1*]   Joan Negre[2, 3]   Jonas Alcaina Mateos[1]   Francesco Carrer[4]

Abstract

The interpretative power of quantitative intra-site spatial analysis has long been recognised by archaeologists. On the contrary, very few ethnoarchaeological works have engaged with the statistical analysis of intra-site pattern of artefacts and ecofacts. Nonetheless, ethnoarchaeology is uniquely placed to guide and assist in the identification of sound techniques and in the development of the methodology of analysis. The possibility to directly connect the distribution of the proxies under study with the activity that generated such distribution, provide the means of fine-tuning the tools we use for identifying these patterns in the archaeological record. As a consequence, archaeological methodology and interpretation both benefit from the application of improved methodologies. In this paper we review some of the most common techniques used to analyse activity areas and we provide examples of the application of such techniques to ethnoarchaeological contexts. The aim is to assess how archaeology can benefit from the intra-site spatial analysis of ethnoarchaeological contexts.

Keywords

Spatial statistics, Geostatistics, Ethnoarchaeology, Quantitative Methods, Activity Areas

[1] CaSEs Research Group, Department of Humanities, Universitat Pompeu Fabra. C/Trias Fargas 25-27, 08005 Barcelona, Spain

[2] Centro Austral de Investigaciones Científicas, Consejo Nacional de Investigaciones Científicas y Técnicas, Bernardo Houssay 200, V9410CAB, Ushuaia (Tierra del Fuego), Argentina

[3] Laboratori d'Arqueologia Quantitativa, Departament de Prehistòria, Universitat Autònoma de Barcelona, 08193 Bellaterra (Cerdanyola del Vallès, Barcelona), Spain

[4] McCord Centre for Landscape, School of Classics, History and Archaeology, Newcastle University, Queen Victoria Road, Newcastle upon Tyne (UK), NE1 7RU

[*] *Corresponding author: carla.lancelotti@upf.edu*

1   Introduction

Intra-site pattern analysis is of crucial importance for archaeologists to unravel past human activities and, as a consequence, past human behaviour. In addition, it can assist in the interpretation of site-formation processes and post-depositional taphonomic processes (Driscoll et al. 2015). Thus, archaeologists have long since recognised the importance of thoroughly analyse spatial patterns and have devoted much effort to the development of sound analytical and interpretative methods. Initially, archaeological research focused on the spatial organisation of artefacts and ecofacts (Schiffer 1987). In the last decades more attention has been given to the microscopic traces of human activities and natural processes at site scale (e.g. Kimura et al 2001, Mallol et al 2007, Pecci et al. 2013). This critical change in perspective was triggered by the development of new analytical methods that allowed the identification of microscopic traces of human activities and provided reliable results, which enabled detailed and significant inferences (for an overview of microarchaeology see Weiner 2010). Regardless of the method applied for the investigation of intra-site contexts, the complexity of most of the patterns highlighted forced archaeologists to create specific interpretative toolbox to link these patterns to spatial behaviours and processes. Ethnoarchaeology is one of the best suited methods for addressing this issue (David and Kramer 2001, 255–283).

Ethnoarchaeology developed within the theoretical framework of New Archaeology, as an approach for correlating material culture and human behaviour (social and cultural) through the construction of law-like relations (Cunningham 2009; Binford and Sabloff 1982; Shiffer 1995). A critique to this vision came from those who conceived ethnoarchaeology as a source of analogy who aimed at creating analogical models to complement and improve archaeological interpretation (David and Kramer 2001; Skibo 2009). Far from being antithetical, these two visions can be combined so that ethnoarchaeology becomes a middle-ground source of model formation whereby archaeologists adapt their research questions to ethnographic methods and contexts and test their theory against material evidence. Furthermore, the potential of ethnoarchaeology as a source of methodological refinement has been somehow overlooked. In this paper we review some of the most common techniques available for intra-site spatial analysis and we offer some examples of advanced quantitative and multi-scalar techniques applied to ethnoarchaeological contexts.

Undoubtedly spatial analysis in ethnoarchaeology has been pioneered by the work of Luis Barba and colleagues who started researching activity areas through the distribution of chemical elements and residues in the late 1970s (Barba and Bello 1978; Barba and Manzanilla 1982; Barba and Denise 1984; Barba 1990, 2007; Barba and Ortiz 1992; Barba et al. 1995). These were amongst the first

works that used quantitative methods and spatial distribution analysis of activity areas in ethnoarchaeological contexts, focussed on microscopic and intangible elements. Further ethnoarchaeological investigation of spatial organization in hunter-gatherer (Binford 1978; O'Connell 1987; Yellen 1977) and pastoral campsites (Gifford and Behrensmeyer 1977; Simms 1988), as well as in permanent households (Deal 1985), fostered the understanding of formation processes. Later, the development of sediment and soil analysis in ethnographic contexts (Brochier 1992) provided archaeology (as well as geoarchaeology and palaeobotany) with reliable and generalizable reference datasets, which promoted the rapid development of this field of research in the recent years (Mallol et al. 2007; Milek 2012; Shahack-Gross et al. 2004; Shahack-Gross, Simons, and Ambrose 2008). Furthermore, ethnoarchaeology triggered new theoretical approaches for the analysis of intra-site patterns, proposed alternative markers for the identification of activities and post-depositional processes and highlighted what can and cannot be identified within an archaeological context (see Yu 2014 for an overview of the impact of ethnoarchaeology on archaeological methods and theory). However, it was rarely used for creating models that enabled the relationship between spatial behaviour and spatial patterns to be unlocked. The main reason for this is often the lack of an appropriate theoretical approach that justify the use of quantitative methods for spatial analysis in ethnoarchaeological research (see Codding and Bird 2015; O,Connell 1995).

1.1   Overview of spatial analysis in ethnoarchaeology

While sophisticated and state-of-the-art spatial analysis started being applied in archaeology since the 1960s (Hodder and Orton 1976; Carr 1984; Djindjian 2009), ethnoarchaeological inferences were still largely based on qualitative, or semiquantitative, estimations and visual assessments of spatial patterns (see for example Binford 2001). Density maps and contour lines were still the main output used to represent spatial aggregation and segregation of features and activities (Yellen 1977; O'Connell 1987; Simms 1988). During the 1980s quantitative methods started being applied in intra-site ethnoarchaeology.

One of the earliest examples is Spurling's and Hayden's research on assemblage patterns within two hunter-gatherer sites of Western Australia (Spurling and Hayden 1984). Due to the limited dataset available, they applied Fisher's Exact test to investigate spatial co-association between artefact/ecofact categories. The frequency of presence and absence of items (based on mean-splitting procedure) per spatial unit, was tested for each pair of item classes against hypothesis of independence. The statistically inferred toolkits were compared with the tool associations derived from the ethnographic data. The results enabled an evaluation of potentials and limitations of statistical tests for the identification of activity areas in archaeological sites.

Roger Cribb studied the distribution of items within and around the domestic areas of nomadic pastoralist campsites in Anatolia (Cribb 1991). Nearest neighbour analysis and bivariate test of association were used to assess the degree of clustering among and between each object category. In addition, Principal Component Analysis (PCA) was used to estimate the co-occurrence of different categories within spatial units. The plot of the PCA scores for each spatial unit enabled the investigation of the degree of spatial coherence for each component. Spatial patterns of objects were identified within the campsites and compared with the location of the actual activity areas. The interpretation of the outcomes provided interesting insights for understanding the spatial organization of assemblages within archaeological nomadic sites.

Despite the promising results of these pioneering studies, later ethnoarchaeological projects kept relying mostly on visual assessment to identify spatial patterns (Bamforth, Becker, and Hudson 2005; Svoboda et al. 2011). Sedimentological studies of ethnoarchaeological contexts, on the other hand, focused on the identification of local markers of activities and depositional processes, rather than on their spatial distribution (Brochier 1992; Shahack-Gross et al. 2004). While quantitative methods became standard tools for the analysis of archaeological sites and landscapes (Orton 2004; Lloyd and Atkinson 2004), ethnoarchaeology was still providing qualitative inferences on the relationship between spatial behaviours and material evidence.

The aim of this paper is to place emphasis on the potentials of ethnoarchaeological research to reinforce analytical methods, offering a way of testing and fine-tuning the tools applied to the study of archaeological contexts and illustrate how archaeology can benefit from a wider and more sophisticated application of quantitative methods in the ethnoarchaeology of intra-site contexts. In order to do so, some of the most common techniques currently used in intra-site archaeological research are detailed and described. Benefits and drawbacks of their use in ethnoarchaeology are pointed out, recent examples of the application of the same methods in ethnoarchaeology are provided and the potentials for archaeological research are evaluated. Further perspectives and pathways of development are proposed in conclusion.

## 2	Spatial statistics and geostatistics: theoretical insights and recent applications

The main goal of spatial analysis is to identify the spatial structure inherent to a certain distribution of points (being these artefacts, ecofacts or structures), in order to facilitate the inference about the specific actions that produced the observed pattern (Barceló and Maximiano 2008). Spatial statistics and geostatistics are the main fields of study that engage with the issues of revealing, predicting and measuring the uncertainty in the prediction of spatial patterns.

### 2.1	Analysis of coordinated data

The study of coordinated remains (i.e. remains with a recorded spatial position) has been a pioneering and ground-breaking field ever since the importance of spatial distributions was recognised by archaeologists (Whallon 1974 and 1975). The latest developments of spatial statistics offer a set of powerful tools to extract as much information as possible from a series of coordinated data. The three main contribution of spatial statistics are: 1) the description of spatial structure; 2) the analysis of clustering; and 3) the testing of spatial variability. These three aspects fall back into the range of influence of Point Pattern Analysis methods, a set of techniques focused on the identification of the main structural characteristics of a point process (Gatrell et al. 1996: 256, Longley 2000, Fotheringham et al. 2000: 15-22).

#### 2.1.1	Spatial descriptive statistics

Spatial descriptive statistics was first standardized in the pioneering works of Whallon, and their subsequent revisions (Whallon 1984). Nowadays, any spatial statistics software, including most of the GIS platforms, provides basic spatial descriptors, such as the standard deviational ellipse or the centroid of the distribution. Particularly important within spatial descriptive statistics is exploratory spatial data analysis, a collection of techniques used to describe and visualise spatial distributions, identify atypical locations or spatial outliers and discover patterns of spatial association (Anselin 1998). One of the most developed descriptive tools, currently still used in a large number of intra-site spatial analyses, are the Kernel algorithms for density estimation (KDE). They provide a smoother representation of the raw data, hence facilitating both the presentation and comparison of several data sets in a single figure (Baxter 1997). One of the most important improvements of these algorithms, especially relevant for archaeology, is the development of an adjusted algorithm, which takes into consideration the presence of barriers (Gribov and Krivoruchko 2011).

A classic case study of the use for bivariate KDEs is the ethnoarchaeological data set from Mask Site (Binford 1978b). The data, as presented by Blankholm (1991), consist of the spatial coordinates of

five classes of find that occur in the archaeological record, such as artefacts, large bones and bone splinters. In this example, the KDE approach allows the identification of a structure in the remains' deposition, as well as providing an informal mean of cluster analysis, indicating a relationship between the artifacts and the hearths. Kernel approaches were also applied in two studies in Tierra del Fuego, where hypotheses based on ethnographic literature (rather than direct observation of living communities) were tested on two archaeological contexts: Tunel VII and Lanashuaia II. In the former this tool was implemented as a way to establish an analytical-interpretive discourse about inside/outside limits of archaeological domestic structures (Maximiano 2012). The latter focused on framing activity areas, discriminating social and natural processes in the formation of the archaeological record by comparison with other ethnoarchaeological contexts located also in the Beagle Channel (Negre et al. 2015).

2.1.2   *Spatial clustering*

In relation to spatial aggregation, there is no doubt that spatial *k*-means clustering (Kintigh and Ammerman 1982) is one of the most popular techniques used by archaeologists in a wide range of intra-site case studies (Blankholm 1991, Savage 1997, Vaquero 1999, Alconini 2004, Lemke 2013). In the last years, other clustering methods have been developed in order to overcome some of the major limitations linked to the *k*-means techniques, such as *k*-medoids, fuzzy clustering, model-based clustering, or density-based clustering algorithms (Baxter 2015). While alternative methods to basic clustering have also shown weaknesses (including problems such as modifiable areal unit, edge effects or smearing), they have shown great potential and are regarded as an encouraging research line (Ducke 2015: 366). An exploratory path to overcome the assumption of surface homogeneity (i.e., movement cost is uniform and symmetric in the whole study-area and it is a function of Euclidean distance between locations) has been recently proposed, modelling a new spatial distribution based on the pairwise cost-based distance matrix of the original observations, in order to represent them in a pseudo-Euclidean framework (Løland and Høst 2003, Negre 2015). In this way, classification techniques can be applied over a spatial distribution of objects re-sorted taking into consideration the presence of barriers, such as the walls of a domestic unit context.

Clustering techniques were among the earliest quantitative methods applied in ethnoarchaeology. Susan A. Gregg, Keith W. Kintigh and Robert Whallon re-used Yellen's ethnoarchaeological data from !Kung campsites (Yellen 1977) to test the validity and robustness of some spatial analytical methods (Gregg, Kintigh, and Whallon 1991). According to Yellen's observations and inferences, the authors expected nuclear areas to be mirrored in discrete clustering of items and similar patterns to be identifiable within each nuclear area. *k*-means was applied to evaluate the spatial distribution of

materials within the campsite; sum-squared-error (SSE) of input data and random runs were plotted against clustering stages, and their difference enabled the identification of significant clustering levels. Spatial patterning in assemblage composition was investigated using unconstrained clustering approach, by interpolating material classes densities from smoothed assemblage density. Clustering (minimization of within-group variance) was estimated from standardized density of each material class, and SSE were plotted against clustering stages. The results conformed with Yellen's interpretation of activity areas distribution within the campsite.

2.1.3   Spatial variability

The re-ordination approach described above can be very useful for other analyses, such as testing spatial variability or comparing patterns against a known model. Ripley's K-function is commonly used to perform these tests, where the null hypothesis is that the point process under consideration follows a homogeneous Poisson distribution and the alternative is that the distribution exhibits clustering or inhibitory behaviour (Ripley 1981: 158-159, Veen and Schoenberg 2006: 293). Other statistical approaches allow to better assess the correlation of the empirical observations with theoretical models formulated through ethnoarchaeological inference (Barceló and Maximiano 2008), such as the bivariate normality to simulate intentionality patterns (Mardia 1985), bivariate exponential distributions to simulate cleaning patterns (Kundu and Gupta 2009, Shoaee and Khorram 2012) or multimodal distributions to simulate social interaction patterns (Eiben and Smith 2003).

The analysis of spatial variability provides the means to discern between regular and stochastic processes by assessing the deviation from spatial homogeneity. In a similar way that Ripley's K-function worked for coordinated remains, global Moran's I index measures similarities between both the location and attribute values of spatial objects (Cliff and Ord 1973, Hodder and Orton 1976). In addition, a significance test can be performed to substantiate the results of the test (Wheatley and Gillings 2002: 119). Once the non-homogeneity and spatial dependency of the analysed data is confirmed, the correlation of the empirical observation of frequencies with the theoretical spatial models can be checked. A set of techniques, which are encompassed under the global term of Local Indicators of Spatial Association (LISA) is available to test for similarities between the models and the observed. It includes, among others, techniques such as Anselin local Moran's I that are fitted to detecting hot spots, for example areas (in terms of probability) where high frequencies of intentionally deposed materials can be highlighted (Anselin 1995).

These methods were recently applied to investigate ethnoarchaeological contexts. Biagetti et al (2016), used spatial statistics to analyse spatial distributions of small pastoral campsites of the Kel Tadrart Tuareg in SW Libya. Both standard statistical techniques and tailored analysis were used to test for the degree of separation between domestic and livestock structures as well as to establish whether spaces dedicated to guests were spatially isolated from the rest of the camp. First, they provided an estimation of the area occupied by the settlement through the application of Ripley-Rasson estimate (Ripley and Rasson 1977). Subsequently they conducted a bivariate *L*-function analysis (Diggle 2003; see also Crema and Bianchi 2013 and Bevan et al. 2013 for archaeological examples of its application) dividing the dataset in domestic and livestock structures and comparing the results of this statistics against a simulated envelop, which resulted in the confirmation of the segregation of these two classes of structures. Finally, they computed the probability that the guest structures held a special position in the campsite (isolated from the rest of the structure) by comparing their actual distance from the nearest structure to the simulated distance (considering any of the structure in the campsite as a guest space independently of its real function). Their results confirmed the hypothesis that such structures are indeed slightly isolated from the rest of the campsite thus stressing the importance of guests and the respect of their privacy, both highly valued by the Tuareg.
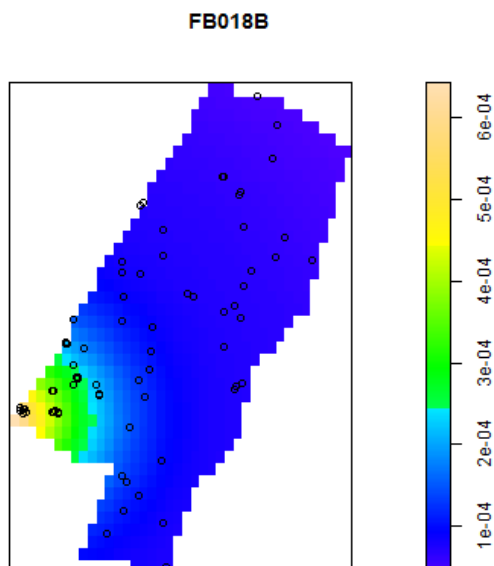
Spatial statistics the analysis of spatial dependency were used to investigate the assemblage pattern within two currently used pastoral huts in the western Italian Alps (Carrer 2015). The purpose was to assess whether spatial organisation of discarded objects enabled the identification of intra-site activity areas in seasonal contexts, and how post-depositional disturbances affected this spatial organization. In order to investigate aggregation and segregation at different scales, homogeneous *L*-Function was estimated for the studied assemblages.

Inhomogenous *L*-Function was used to assess the existence of first order effects influencing the spatial variation of intensity of the point process. Participant observation suggested that most of the activities in one of the huts were carried out around the fireplace, and this behaviour was mirrored

in the higher density of items around the fireplace area. Intra-site trampling was ethnographically recorded as well, and its consequence for the assemblage were investigated using autocorrelation analysis. The hypothesis was that fragmented objects were closer together in the most trampled sectors of the huts whereas bulky objects were displaced in the peripheries. Moran's *I* correlogram and omnidirectional variogram were used to identify autocorrelation between the location of the object and their size; directional variogram was applied to assess the directionality of trampling. This ethnoarchaeological case-study enhanced the reliability of assemblage pattern analysis in seasonal archaeological contexts, suggesting how the intensity of occupation rather than the duration had a key-role in the spatial organisation of activity areas. It also contributed to highlight the importance of autocorrelation analysis for disentangling the consequences of trampling in the spatial organisation of archaeological assemblages.

2.2   *Geostatistics*

Geostatistics is mainly concerned with prediction of data points, modelling a stochastic process over a continuous spatial region from a partial group of observations. It is based on the main assumption that neighboring points have similar values and the strength of their correlation decreases when their distance increases. It was firstly developed by D.G. Krige (1951) and G. Matheron (1963) in the central decades of the twentieth century, to predict mining locations. Recently, Lloyd and Atkinson (2004) provided an overview of the potential of geostatistics in archaeology, and stressed the importance of using these techniques where the spatial variation of archaeological variables is under study.

Geostatistics, being a model-based interpolation technique provides the possibility of performing model-selection prior to interpolation thus increasing the reliability of the new data produced. In addition, through techniques such as leave-one-out cross-validation, it is possible to test for the reliability of the predictions thus giving a quantification of the uncertainty inherent to the data. This has great advantages over deterministic interpolation methods, such as inverse distance weighting (IDW), bilinear interpolation or nearest-neighbor interpolation, in that it consider the value to be predicted as defined by a cumulative distribution function (Cressi 1993). One of the most commonly used geostatistical method is kriging, which in its basic form (Ordinary Kriging) uses optimized weights of neighbouring samples according to a mathematical model and a system of equations. The semi-variogram model is not used only to optimize the samples' weights but also to provide useful information on the spatial structure of the points. Co-kriging, a variation of the method, exploits the relationships between the variable to be predicted and a more-intensely sampled variable to calculate more accurate and more precise predictions (Chun and Griffith 2013). More recently,

2

geostatistical interpolation has been further improved by the application of advanced simulation techniques such as Markov Random Field models (Besag 1974), Multiple-point geostatistics (Guardiano and Srivastava 1993) and pseudo-genetic models (Behzadi and Alesheikh 2008). These techniques increase the precision of interpolated values, especially when distance lag between sampled points is high or when the initial dataset is small (few observation points).

Geostatistics has been recently applied to ethnoarchaeological contexts within the framework of research on the definition of anthropic activity markers (AAMs) (Rondelli et al 2014). AAMs are defined as models, driven by ethnoarchaeological and experimental archaeological research, and produced by a multi-proxy and multi-scalar spatial analysis of evidences of anthropic activities. The possibility to identify and connect these evidences to the activity that generated the record is pivotal to our understanding of past human behaviour. In their work Rondelli et al., performed a multi-proxy analysis of a traditional house within a farming compound in North Gujarat (India). The floors of the house were firstly investigated through micromorphology thin section to understand their structure and formation. Subsequently, multi-element ICP-AES and spot tests (i.e. protein, fatty acids and phosphates residues) were performed on samples collected in a rectangular lattice of 100 cm. Several interpolation methods were tested on both single elements, combination of elements (sums) and PCAs. Although all methods display similar trends in most cases -most notably, the division between the internal space and the veranda- only the maps created with Regression Kriging (RK) highlight clearly the separation between the two areas. Similarly, RK represents at best the activity areas and the division of domestic space as identified through the ethnographic survey and interviews.

A further development of this work has dealt with the issue of considering physical barriers during interpolation (e.g. walls) by implementing the use of cost-based distances to quantify the correlation between sampling locations (Negre et al., 2016). For testing purposes, the chemical spatial distribution map of Calcium residues was modelled considering the walls of the house as non-transparent barriers affecting its distribution. The most important aspect of this work, therefore, lies on the general methodology for overcoming the geostatistical restriction on the homogeneity of the prediction region. In the study of geochemical signatures on archaeological floors, that limitation is a major problem for obtaining significant interpretation of the sampling data. In summary, taking into account the heterogeneity of the working region using the cost-based distance led to important differences in predictions. In some cases, this variability reached values near 1% between both kriging predictors, which may account for the benefits of the cost-based method. Archaeological reading of chemical signatures in domestic floors is highly related to the combination of diverse

element residues, such as different combinations of Potassium, Calcium, Phosphorus, Magnesium and Strontium for identifying deposited food remains, living room sediments, enclosed spaces or burning areas (Rondelli et al. 2014: 488). In the presence of several geochemical indicators related at the same time, a perfect delimitation of their distribution areas is crucial for a valid inference.
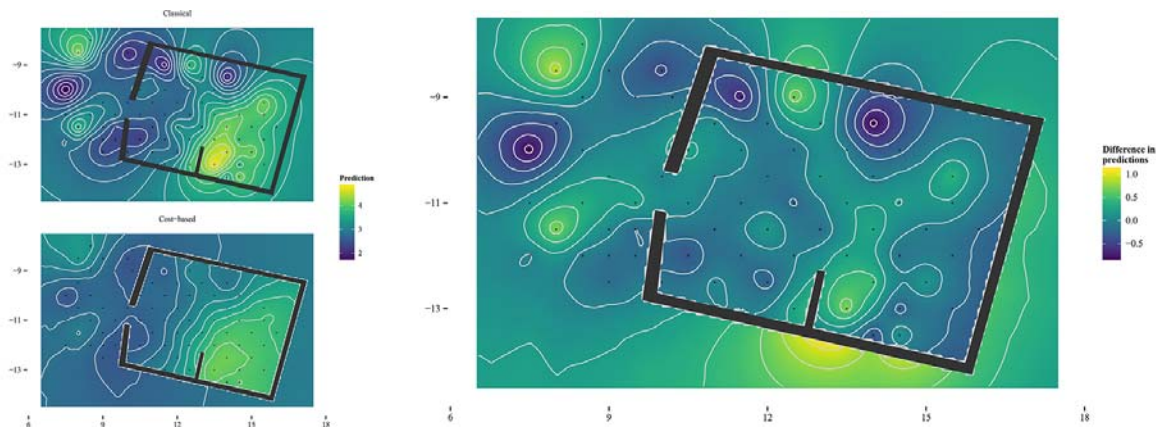
3   Future perspectives

At present, spatial analysis of ethnoarchaeological contexts is still a poorly explored field. If we agree that 'strength lies in numbers' widening the number of case-studies is paramount to push forward the power of ethnoarchaeological analysis. In order for the discipline to grow, we need to be able to apply these methods not only to a greater number of case-studies, but also to increase the time-depth of our investigations by re-visiting the same locations for multiple consecutive occasions within and between years. Indeed, increasing the 'sample size' systematically might be a way to increase precision when estimating unknown parameters, and to find generalised patterns. However, some studies seem to suggest that it is not always the case (see for example, Crowther 1997; Haslam and Tibett 2004; Wells 2010) and that the relationship between sampling accuracy and precision might be more complex.

A second major issue on the ethnoarchaeological research agenda is that of establishing models of the use of space that can be translated to the archaeological context. In this respect ethnoarchaeology is probably the only discipline that can assist in the definition of methodological standard for archaeological research. The direct observation and description of the activities that created the ethnoarchaeological record enable the methodology of analysis to be tuned and moulded to the research question, up to the point of obtaining the desired result. Indeed, by knowing what is the result we expect from the analysis (i.e. the correct representation of the spatial activity observed) we can adjust the statistical and mathematical methods of analysis so that the results we obtain mirror what we witness in the field. Obviously, in order for this to be significant,

we reinstate that it is fundamental to have a good sample size so that we can escape the idiosyncrasy of the single case-study and the biases that can be thus generated.

Lastly, one of the major contribution that ethnoarchaeology can offer to the study of past spatial activities is to provide models for assessing and quantifying taphonomical processes that can affect the record. Gregg et al. (1991) conducted a pioneer study in this perspective, quantitatively analysing ethnoarchaeological data and then simulating the decay of organic material and the movement of small artefacts to reproduce possible archaeological site formation processes. They then compared the results of the analysis conducted on the simulated distributions with the original ethnographic distributions and concluded that, notwithstanding some loss of resolution, the original patterns of human activities were maintained in the simulated archaeological contexts. More recently, Driscoll et al (2015) conducted experimental work on the spatial analysis of trampled flint and Ullah et al (2015) used simulations of micro-refuse distributions to simulate site-formation processes.

A challenge of spatial analysis in ethnoarchaeology lies in the correct correlation of quantitatively recorded spatial data with spatial behaviours. In order to provide a more reliable baseline for analogical reasoning, activity areas and post-depositional processes should be recorded using standardized procedures, rather than observed and sketched. An innovative development in this direction is represented by the work of T.A. Surovell and M. O'Brien with the Dukha Ethnoarchaeological project (https://sites.google.com/site/dukhaethnoarch/home). Their innovative idea concerns the direct tracking of people movements while performing their daily activities rather than analysing the traces these activities leave on the ground. Through the use of photographs taken at regular intervals from a camera set up outside nomadic pastoral campsites in Mongolia, they are able to track people and recording not only their use of space but also the age and sex of the performers, the duration of the activities and the environmental conditions in which these activities take place (Surovell and O'Brien,2016). This type of work, coupled with the analysis of material traces (macro- and microscopic) left behind when camps are dismounted, can possibly represent one of the most powerful developments of spatial analysis in ethnoarchaeological contexts.

Finally, ethnoarchaeology offers the possibility of addressing one of the main assumptions of spatial statistics: stationarity. According to this principle, the spatial correlation between any two points is a function of their Euclidean distance and does not depend on their exact location. Formally, a constant intensity and a uniform correlation depending purely on the lag vector between a pair of points, can be assumed for most spatial point processes (Møller and Toftaker 2012). This framework has been vastly used to analyse all kind of intra-site spatial patterns, surmising that the activities

took place on a homogeneous and undifferentiated surface. When confronted with a general archaeological set of items, distributed over a regular surface without boundaries, it is sensible to think of Euclidean distance as the natural argument for a correlation function. However, the presence of total or partial barriers, such as the walls of a domestic unit or even more subtle irregularities of the floor, the presence of spatial attractors or a constraining morphology might affect this correlation (López-Quílez and Muñoz 2009).

The physical reality in which human activities took place in the past is often more complex than a geometric container and for this reason is too simple to generalise spatial phenomena with Euclidean metrics (Kitchin 2009). Alternative measures to Euclidean distances have been tested, for example, to understand the optimal paths between settlements (Smith 1989, Marble 1996, Llobera 2000, Bell and Lock 2000, Wheatley and Gillings 2002: 155-156, van Leusen 2002, Conolly and Lake 2006: 221-224). Nonetheless, the implementation of cost-based distances in spatial statistics and geostatistical applications in archaeology is far from being up to date. New approaches have been recently proposed in order to overcome this important limitation in archaeological case studies, through the application of cost-based kriging to an ethnoarchaeological case-study (Negre and Muñoz 2015). In this work, the results published by Rondelli et al. (2014) were compared with a newly developed bespoke method that takes into account physical barriers in the spatial distribution of elements. Although the results are not extremely different between the two methods, the cost-distance model corrects many small discrepancies in the process of modelling the distribution surfaces of the elements under study. This is fundamental for archaeological cases where data are less abundant than in ethnography and the summed errors might be greater.

4   Summary and conclusions

Spatial statistical analysis and geostatistics are powerful tools that archaeologists have been employing to study the remains of human activities in the past and interpret the behaviour that generated those remains. On the contrary, ethnoarchaeologists have so far concentrated more to the qualitative recording and analysis of spatial patterns. Ethnoarchaeology is uniquely placed to provide the missing link between specific human actions and what remains of those actions on the ground. Within this privileged contexts, quantitative spatial analysis offers the opportunity to refine our methodology and techniques. As a consequence, it delivers meaningful models of activity markers that can be confidently applied to study archaeological contexts.

Several authors see the archaeological record as a palimpsest of activities, where all but the remains of the latest activity performed can be recovered (for a review of this concept see Bailey 2007).

However, in several contexts and depending on the substrate on which such activities are performed, we are not looking at a true palimpsest -- i.e. something where all traces of previous activities have been completely erased and only the latest can be seen-- but at the combined results of multiple, repeated activities. This is the case of residue of domestic activities performed on a mud and clay surface, such as in the work of Rondelli et al. (2014) and Negre et al. (2016) referred to previously. Several ethnoarchaeological studies tackle (though indirectly) the problem of archaeology as a palimpsest and demonstrate that, with the proper methodology and the application of several parallel lines of investigations, this can be resolved (see for example, Shahack-Gross et al. 2004; Hutson and Terry 2006; Hutson et al. 2007; Tsartsidou et al. 2009; Friesem et al. 2013, 2014;). In this situation, a meaningful way of unfolding this superimposition of activities is to directly investigate the activities, and their material remains, through the application of quantitative techniques that can help in assessing the input of synchronous and diachronic depositions. Another way of assessing the impact of time over the assemblage is provided by the study performed by Carrer (2015) on Alpine huts. In this case, spatial statistics combined with participative interviews, allowed the author to conclude that the intensity of occupation has a much stronger effect than time on the spatial deposition of artefacts.

Ethnoarchaeology also offers the opportunity to formulate and test archaeological hypotheses on human activities. As presented above, Negre and Muñoz (2015) developed a novel cost-based kriging prediction method aimed at considering the importance of physical barriers in the distribution of activity proxies, induced by the information gathered ethnographically. Later they applied and compared their method on an ethnoarchaeological case-study so to test the robustness of their method and its applicability in archaeological cases. Similarly, spatial statistics on modern Tuareg campsites allowed Biagetti et al. (2016) to quantitatively test the degree of structure separation, which had been previously proposed for pastoral sites on the basis of ethnographic interviews. The possibility of quantifying and thus normalizing this reported behaviour implies its possible recognition in archaeological contexts. Notwithstanding the taphonomic problems related to the identification of structures in nomadic campsites, which is out of the scope of this paper, the work of Surovell and O'Brien (2016) presents a step forward in linking actions and remains. Indeed, through the direct recording of the people's movements, they reinforce the interview stage, where pieces of information can be missed or misinterpreted.

The case-studies presented in this paper contribute to showcase the methodological and interpretative potentials of quantitative spatial analysis in ethnoarchaeological research. Lloyd and Atkinson (2004) pinpoint the supposed complexity of geostatistical techniques as a major obstacle in

their widespread use in archaeology. The lack of quantitative training in most humanistic degrees might indeed contribute to the feeling of being inadequately prepared to tackle this challenge. However, a series of tools exist that allow a simple and straightforward application of statistical techniques (e.g. user-friendly software, online forums and debates, free online courses, etc.). In perspective, we maintain that open data and open software policies are absolutely fundamental to help users overcoming the barriers of applying tools that they feel unfamiliar with (Lake 2012). Therefore we support, as shown in the case-studies detailed in this paper, the release of datasets and codes used for analysis, to be openly reused, modified and improved by other scholars. We hope that this will encourage people to address this type of issues with less diffidence and more studies of this kind will become available in the near future.

Bibliography

Alconini, S. 2004. The southeastern Inka frontier against the Chiriguanos: structure and dynamics of the Inka imperial borderlands. *Latin American Antiquity* **15**, 389-418.

Anselin, L. 1995. Local Indicators of Spatial Association (LISA). *Geographical Analysis* **27**(2), 93-115.

Anselin, L. 1998. Exploratory spatial data analysis in a geocomputational environment, pp. 77-94 in Longley, P. A., Brooks, S. M., McDonnell, R. and Macmillian W. (eds.). *Geocomputation: a primer*. New York: Wiley and Sons.

Bailey, G. 2007. Time perspectives, palimpsests and the archaeology of time. *Journal of Anthropological Archaeology* **26**(2), 198-223.

Bamforth, D. B., Becker, M., and Hudson, J. 2005. Intrasite Spatial Analysis, Ethnoarchaeology, and Paleoindian Land-Use on the Great Plains: The Allen Site. *American Antiquity* **70**(3), 561–80.

Barba, L. and Bello, G., 1978. Análisis de fosfatos en el piso de una casa habitada actualmente. *Notas Antropológicas* **24**, 188-193.

Barba, L. and Denise P. 1984. Actividades humanas y análisis químico de los suelos: El caso de Osumacinta Viejo, Chiapas. Memorias de la XVII Mesa Redonda de la Sociedad Mexicana de Antropología **II**, 263–277.

Barba, L. 1990. El análisis químico de pisos de unidades habitacionales para determinar sus áreas de actividad, pp. 177–200 in Sugiura, Y. and Serra Puche, M. C. (eds.) Etnoarqueología Coloquio Bosch Gimpera 1988. Mexico City: Instituto de Investigaciones Antropológicas, Universidad Nacional Autónoma de México.

Barba, L., and Ortiz Butrón, A., 1992, Análisis químico de pisos de ocupación: un caso etnográfico en Tlaxcala, México. *Latin American Antiquity* **3**, 63–82.

Barba, L. 2007. Chemical residues in lime-plastered archaeological floors. *Geoarchaeology*, **22**(4), 439-452.

Barceló, J. A and Maximiano, A. 2008. Some notes regarding distributional analysis of spatial data, pp. 1-6 in Posluschny, A., Lambers, K. and Herzog, I. (eds.). *Proceedings of the 35th International Conference on Computer Applications and Quantitative Methods in Archaeology (CAA)*, Bonn: Rudolf Habelt GmbH.

Baxter, M. J. 1997. Some archaeological applications of kernel density estimates. *Journal of Archaeological Science* **24**, 347-354.

Baxter, M. J. 2015. Spatial *k*-means clustering in archaeology - variations on a theme. Working paper. https://www.academia.edu/18142974/ (03/12/2015)

Bell, T. and G. Lock. 2000. Topographic and cultural influences on walking the Ridgeway in later prehistoric times, pp. 85–100 in Lock, G. (ed.). *Beyond the Map: Archaeology and Spatial Technologies*. Amsterdam: IOS Press.

Besag, J. 1974. Spatial interactions and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)* **36**(2), 192-236.

Bevan, A., Crema, E. R., Li, X. and Palmisano, A. 2013. Intensities, interactions and uncertainties: Some new approaches to archaeological distributions, pp 27-52 in Bevan, A., and Lake, M. (eds.), *Computational Approaches to Archaeological Space,* Walnut Creek: Left Coast Press.

Behzadi, S. and Alesheikh, A. A. 2008. A pseudo genetic algorithm to solve best path problem. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **37**(B2), 253-256.

Blankholm, H. P. 1991. *Intrasite Spatial Analysis in Theory and Practice*. Aarhus: Aarhus University Press.

Biagetti, S., Alcaina-Mateos, J. and Crema E. 2016. A matter of ephemerality. The study of Kel Tadrart Tuareg (SW Libya) campsites via quantitative spatial analysis. *Ecology and Society* **21**(1), 42

Binford, L. 1978. Dimensional Analysis of Behavior and Site Structure: Learning from an Eskimo Hunting Stand. *American Antiquity* **43**(3), 330–61.

Binford, L. 2001. *Constructing frames of reference.* Oakland, CA: University of California Press.

Binford, L. R. and Sabloff, J. A. 1982. Paradigms, systematics, and archaeology. *Journal of Anthropological Research* **38**, 137–153.

Brochier, J. E. 1992. Shepherds and Sediments: Geo-Ethnoarchaeology of Pastoral Sites. *Journal of Anthropological Archaeology* **11**, 47–102.

Carr, C. 1984. The Nature of Organization of Intrasite Archaeological Records and Spatial Analytic Approaches to Their Investigation. *Advances in Archaeological Method and Theory* **7**, 103–222.

Carrer, F. 2015. Interpreting Intra-Site Spatial Patterns in Seasonal Contexts: An Ethnoarchaeological Case-Study from the Western Alps. *Journal of Archaeological Method and Theory*. doi:10.1007/s10816-015-9268-5

Chun, Y. and Griffith D. A. 2013. Spatial Statistics and Geostatistics. Los Angeles|London|New Delhi|Singapore|Washington DC: SAGE Publications.

Cliff, A. D. and Ord, J. K. 1973. *Spatial autocorrelation*. London: Pion.

Codding, B.F., and Bird, D.W. 2015. Behavioral ecology and the future of archaeological science. *Journal of Archaeological Science* **56**, 9-20.

Conolly, J. and Lake, M. 2006. *Geographical Information Systems in Archaeology*. Cambridge: Cambridge University Press.

Crema, E. R., and Bianchi, E. 2013. Looking for patterns in the noise: non-site spatial-analysis in Sebkha Kelbia, pp. 385-395 in Mulazzani, S. (ed.), *Le Capsien de hergla (Tunisie). Culture, environnement et économie*. Frankfurt: Africa Magna Verlag.

Cressie, N.A.C. 1993. *Statistics for Spatial Data*. Wiley Series in Probability and Statistics. Wiley-Interscience: New York.

Cribb, R. 1991. *Nomads in Archaeology*. Cambridge: Cambridge University Press.

Cunningham, J. 2009. Ethnoarchaeology beyond correlates. *Ethnoarchaeology* **1**, 115-136.

David, N. and C. Kramer. 2001. *Ethnoarchaeology in Action*. Cambridge: Cambridge University Press.

Deal, M. 1985. Household Pottery Disposal in the Maya Highlands: An Ethnoarchaeological Interpretation. *Journal of Anthropological Archaeology* **4**, 243–91.

Diggle, P. J. 2003. *Statistical Analysis of Spatial Point Patterns (second edition)*. London: Edward Arnold.

Djindjian, F. 2009. The Golden Years for Mathematics and Computers in Archaeology (1965-1985). *Archeologia E Calcolatori* **20**: 61–73.

Driscoll, K., Alcaina, J., Éguez, N., Mangado, X., Fullola, J.-M. and Tejero, J.-M. 2015. Trampled under foot: A quartz and chert human trampling experiment at the Cova del Parco rock shelter, Spain. *Quaternary International*. doi:10.1016/j.quaint.2015.04.054.

Ducke, B. 2015. Spatial cluster detection in Archaeology: current theory and practice, pp. 353-368 in: Barceló, J. A. and Bogdanovic, I. (eds.), *Mathematics and Archaeology*. Boca Raton: CRC Press.

Eiben, A. E. and Smith, J. E. 2003. Multimodal problems and spatial distribution, pp. 153-172 in Eiben, A. E. and Smith, J. E (eds.), *Introduction to Evolutionary Computing*. Berlin: Springer.

Fotheringham, A. S., Brunsdon, C. and Charlton, M. 2000. *Quantitative geography: perspective on spatial data analysis*. London: SAGE.

Friesem, D., Tsartsidou, G., Karkanas, P. and Shahack-Gross R. 2013. Where are the roofs? A geo-ethnoarchaeological study of mud brick structures and their collapse processes, focusing on the identification of roofs. *Archaeological and Anthropological Science* **6**, 73-92.

Friesem, D., Karkanas, P., Tsartsidou, G. and Shahack-Gross R. 2014. Sedimentary processes involved in mud brick degradation in temperate environments: a micromorphological approach in an ethnoarchaeological context in northern Greece. *Journal of Archaeological Science* **41**, 556-567.

Gatrell, A. C., Bailey, T. C., Diggle, P. J. and Rowlingson, B. S. 1996. Spatial point pattern analysis and its application in geographical epidemiology. *Transactions of the Institute of British Geographers* **21**(1), 256-274.

Gifford, D. P., and Behrensmeyer, A. K. . 1977. Observed Formation and Burial of a Recent Human Occupation Site in Kenya. *Quaternary International: The Journal of the International Union for Quaternary Research* **8**(3), 245–56.

Gregg, S. A., Kintigh, K. W. and Whallon, R. 1991. Linking Ethnoarchaeological Interpretation and Archaeological Data, pp. 149–96 in Kroll, E. M. and Price T. D. (eds.) *The Interpretation of Archaeological Spatial Patterning*. New York: Springer.

Gribov, A. and Krivoruchko, K. 2011. Local polynomials for data detrending and interpolation in the presence of barriers. *Stochastic Environmental Research and Risk Assessment* **25**(8), 1057-1063.

Guardiano, F. and Srivastava, M. 1993. Multivariate geostatistics: Beyond bivariate moments, pp. 133–144 in Soares, A. (ed) *Geostatistics-Troiá*. Dordrecht, Netherlands: Kluwier Acad.

Hodder, I. and Orton, C. 1976. *Spatial Analysis in Archaeology*. Cambridge: Cambridge University Press.

Hutson, S. R., Stanton, T. W., Magnoni, A., Terry, R. E., Craner, J. 2007. Beyond the buildings: formation processes of ancient Maya houselots and methods for the study of non architectural space. *Journal of Anthropological Archaeology* **26**, 442–473.

Hutson, S. R. and Terry, R. E. 2006. Recovering social and cultural dynamics from plaster floors: chemical analyses at ancient Chunchucmil, Yucatan, Mexico. *Journal of Archaeological Science* **33**, 391–404.

Kimura, B., Brandt, S. A., Hardy B. L. and Hauswirth, W. W. 2001. Analysis of DNA from ethnoarchaeological stone scraper. *Journal of Archaeological Science* **28**, 45–53.

Kintigh, K. W. and Ammerman, A. J. 1982. Heuristic approaches to spatial analysis in Archaeology. *American Antiquity* **47**, 31-63.

Kitchin, R. 2009. Space II, pp. 268-275 in Kitchin, R. and Thrift, N. (eds.). *International Encyclopedia of Human Geography*. Oxford: Elsevier.

Krige, D.G. 1951. A statistical approach to some basic mine valuation problems on the Witwatersrand. *Journal of the Chemical, Metallurgical and Mining Society of South Africa* **52**(6), 119-139.

Kunda, D. and Gupta, R. 2009. Bivariate generalized exponential distribution. *Journal of Multivariate Analysis* **100**(4), 581-593.

Lake, M. 2012. Open archaeology. *World Archaeology* **44**(4), 471-478.

Lemke, A .K. 2013. Cutmark systematics: analyzing morphometrics and spatial patterning at Palangana. *Journal of Anthropological Archaeology* **32**, 16-27.

Llobera, M. 2000. Understanding movement: a pilot model towards the sociology of movement, pp. 65–84 in Lock, G. (ed.). *Beyond the Map: Archaeology and Spatial Technologies*. Amsterdam: IOS Press.

Lloyd, C. D. and Atkinson, P. M. 2004. Archaeology and Geostatistics. *Journal of Archaeological Science* **31**, 151–65.

Løland, A. and Høst, G. 2003. Spatial covariance modelling in a complex coastal domain by multidimensional scaling. *Environmetrics* **14**(3), 307-321.

Longley, P. 2000. Spatial analysis in the new Millennium. *Annals of the Association of American Geographers* **90**(1), 157-165.

López-Quílez, A. and Muñoz, F. 2009. Geostatistical computing of acoustic maps in the presence of barriers. *Mathematical and Computer Modelling* **50** (5-6), 929-938.

Mallol, C., Marlowe, F. W., Wood, B. M. and Porter, C. C. 2007. Earth, Wind, and Fire: Ethnoarchaeological Signals of Hadza Fires. *Journal of Archaeological Science* **34**, 2035–52.

Marble, D.F. 1996. T*he human effort involved in movement over natural terrain: a working bibliography*. Technical Report, Ohio State University, Department of Geography.

Mardia, K. 1985. Tests of univariate and multivariate normality, pp. 279-320 in Krishnaiah, P. R. (ed.), *Handbook of Statistics*. Amsterdam: North Holland.

Matheron, G. 1963. Principles of geostatistics, *Economic Geology* **58**(8), 1246-1266.

Maximiano, A. 2012. Geoestadística y arqueología: una nueva perspectiva analítico-interpretativa en el análisis espacial intra-site. *Analitika, Revista de análisis estadístico* **4**(2), 83-95.

Milek, K. B. 2012. Floor Formation Processes and the Interpretation of Site Activity Areas: An Ethnoarchaeological Study of Turf Buildings at Thverá, Northeast Iceland. *Journal of Anthropological Archaeology* **31**, 119–37.

Møller, J. and Toftaker, H. 2012. *Geometric anisotropic spatial point pattern analysis and Cox processes*. Research Report Series, Department of Mathematical Sciences, Aalborg University.

Negre, J. 2015. Non-euclidean distances in Point Pattern Analysis: Anisotropic measures for the study of settlement networks in heterogeneous regions, pp. 369-382 in Barceló, J. A. and Bogdanovic, I. (eds.), *Mathematics and Archaeology*. Boca Ratón: CRC Press.

Negre, J., Muñoz, F. and Lancelotti, C. 2016. Geostatistical modelling of chemical residues on archaeological floors in the presence of barriers. *Journal of Archaeological Science* **70**, 91-101.

Negre, J., Álvarez, M., Briz, I. and Lacrouts, A. 2015. Estadística espacial aplicada al estudio de los procesos de formación del registro arqueológico en sociedades cazadoras-recolectoras fueguinas. Zenodo. https://zenodo.org/record/34594 DOI: 10.5281/zenodo.34594

O'Connell, J. F. 1987. Alyaware Site Structure and Its Archaeological Implications. *American Antiquity* **52**(1), 74–108.

O'Connell, J.F. 1995. Ethnoarchaeology Needs a General Theory of Behavior. *Journal of Archaeological Research* **3**(3), 205-255.

Orton, C. 2004. Point Pattern Revisited. *Archeologia E Calcolatori* **15**, 299–315.

Pecci, A., Cau Ontiveros, M. Á., Valdambrini, and Inserra, F.. 2013. Understanding residues of oil production: chemical analyses of floors in traditional mills. *Journal of Archaeological Science* **40**(2), 883-893.

Ripley, B. 1981. *Spatial statistics.* New York: Wiley.

Ripley, B., and Rasson, J. 1977. Finding the edge of a Poisson forest. *Journal of Applied Probability* **14**, 483–491.

Rondelli, B., Lancelotti, C., Madella, M., Pecci, A., Balbo, A., Perez, J. R., Inserra, F., Gadekar, C., Ontiveros Cau, M. À., Ajithprasad, P. 2014. Anthropic activity markers and spatial variability: an ethnoarchaeological experiment in a domestic unit of Northern Gujarat (India). Journal of Archaeological Science **41**, 482-492.

Savage, S. H. 1997. Descent group competition and economic strategies in predynastic Egypt. *Journal of Anthropological Archaeology* **16**, 226-268.

Schiffer, M. B. 1983. Toward the identification of formation processes. *American Antiquity* **48**(4), 675-706.

Schiffer, M. B. 1987. *Formation Processes of Archaeological Record*. Albuquerque: New Mexico Press.

Schiffer, M, B. 1995. *Behavioral archaeology: First principles*. Salt Lake City: University of Utah Press.

Shahack-Gross, R., Marshall, F., Ryan, K. and Weiner, S. 2004. Reconstruction of Spatial Organization in Abandoned Maasai Settlements: Implications for Site Structure in the Pastoral Neolithic of East Africa. *Journal of Archaeological Science* **31**(10), 1395–1411.

Shahack-Gross, R., Simons, A. and Ambrose, S. H. 2008. Identification of Pastoral Sites Using Stable Nitrogen and Carbon Isotopes from Bulk Sediment Samples: A Case Study in Modern and Archaeological Pastoral Settlements in Kenya. *Journal of Archaeological Science* **35**, 983–90.

Shoaee, Sh. and Khorram, E. 2012. A new absolute continuous bivariate generalized exponential distribution. *Journal of Statistical Planning and Inference* **142**(7), 2203-2220.

Simms, S. R. 1988. The Archaeological Structure of a Bedouin Camp. *Journal of Archaeological Science* **15**, 197–211.

Skibo, J. M. 2009. Archaeological Theory and Snake-Oil Peddling. The Role of Ethnoarchaeology in Archaeology. *Ethnoarchaeology* **1**(1), 27–56.

Smith, T. E. 1989. Shortest-path distances, an axiomatic approach. *Geographical Analysis* **21**(1), 1-31.

Spurling, B., and Hayden, B. 1984. Ethnoarchaeology and Intrasite Spatial Analysis: A Case Study from the Australian West Desert, pp. 224–241 in Hietala, H. J. (ed.) *Intrasite Spatial Analysis in Archaeology*. Cambridge: Cambridge University Press.

Surovell, T.A., O'Brien, M. 2016. Mobility at the Scale of Meters. *Evolutionary Anthropology* **25**(3), 142-152.

Svoboda, J., Sázelová, S., Kosintsev, P. A., Jankovská, V. and Holub, M. 2011. Resources and Spatial Analysis at Actual Nenets Campsites: Ethnoarchaeological Implications. *Journal of Anthropological Archaeology* **30**, 30–43.

Tobler, W. 1970. A computer movie simulating urban growth in the Detroit region. *Economic Geography* **46**, 234-240.

Ullah, I.I., Duffy, P.R. and Banning E.B. 2015. Modernizing Spatial Micro-Refuse Analysis: New Methods for Collecting, Analyzing, and Interpreting the Spatial Patterning of Micro-Refuse from House-Floor Contexts. *Journal of Archaeological Method and Theory* **22**, 1238-1262.

van Leusen, P.M. 2002. *Pattern to process: methodological investigations into the formation and interpretation of spatial patterns in archaeological landscapes*. PhD Thesis, University of Groningen, Groningen.

Vaquero, M. 1999. Intrasite spatial organization of lithic production in the Middle Paleolithic: the evidence of the Abric Romaní (Capellades, Spain). *Antiquity* **73**(281), 493-504.

Veen, A. and Schoenberg, P. 2006. Assessing spatial point process models using weighted *k*-functions: analysis of California earthquakes. *Lecture Notes in Statistics* **185**, 293-306.

Weiner, S. 2010. *Microarchaeology. Beyond the Visible Archaeological Record*. Cambridge: Cambridge University Press.

Whallon, R. 1974. Spatial analysis of occupation floors. I. Applications of dimensional analysis of variance. *American Antiquity* **38**, 320-328.

Whallon, R. 1975. Spatial analysis of occupation floors. II. The application of nearest neighbor analysis. *American Antiquity* **39**, 16-34.

Whallon, R. 1984. Unconstrained clustering for the analysis of spatial distributions in archaeology, pp. 242-276 in Hietala, H. (ed.), *Intrasite spatial analysis in archaeology*. Cambridge: Cambridge University Press.

Wheatley, D. and Gillings, M. 2002. *Spatial technology and Archaeology. The archaeological applications of GIS*. London: Taylor & Francis.

Yellen, J. 1977. *Archaeological Approaches to the Present: Models for Reconstructing the Past*. New York: Academic Press.

Yu, P. L. 2014. Ethnoarchaeology: Building frames of reference for research, pp. 2535-2545 in Smith, C. (ed.), *Encyclopedia of Global Archaeology*, Springer.

Glossary of the principal statistical terms used

Bivariate normality

The theoretical bivariate normal distribution can be used to represent the formation of spatially regular modalities of social actions. It is based on the first Tobler's law of Geography, stating that near things appear to be more related than distant things. If the spatial modality of the social action performed in the past was regular, and its material consequences have not been altered in a significant way by post-depositional processes, the spatial distribution will fit the bivariate normal distribution. Bivariate normal distributions offer a reference model to test the degree of regularity and hence the spatial intentionality of social action.

Centroid of the Distribution

The centroid or mean centre of a spatial distribution is the average *x* and *y* coordinate of all the entities in the study area. It's useful for tracking changes in the distribution through a third vector (e.g. time) or for comparing the distribution of several groups of points in the same area.

Clustering Techniques (*k*-means, *k*-medoids, fuzzy, model-based)

Clustering techniques aim at grouping objects so that the elements of each one group are more similar to each other than to any element outside the group. Spatial k-means clustering is a standard technique often used in archaeological applications for identifying patterns in point scatters. The method is based on the formation of clusters by associating *k* number of centroids with the closest points. Centroids are recalculated and points re-assigned to clusters (when a new centroid is closer) until no further change occurs. An alternative to this algorithm is, for example, the *k*-medoids method, which replaces centroids for medoids. Medoids for clusters are analogous to the median for one-dimensional data. The main advantage of this technique is represented by its lower sensitivity to outliers. The fuzzy clustering methods, instead, enables partial membership of each entity to each cluster. Other methodologies, and in particular model-based methods, depend on the definition of a probabilistic model of clustering.

Correlogram

A graphics that shows correlation between two variables. A typical example is a time series where values of one variable (e.g. temperature) are displayed according to time points.

Cumulative distribution function

In statistics, a cumulative distribution function describes the probability that a random variable (which follows a specific distribution) will take a value less than or equal to a specific number. In geostatistics, the total variance of a predicted estimation is described by a cumulative distribution, thus providing a measure of the uncertainty (in contrast to other deterministic methods).

Fisher's Exact test

Fisher's Exact test is a statistical test used to assess the significance of association between two classes of categorical data. It is one of several exact test insofar as the significance of the deviation from the null-hypothesis is given as an exact value even for small sample-size. This test can be used to assess the significance of spatial correlation between two categories of artifact/ecofacts.

Homogeneous & Inhomogeneous Poisson Process

A Poisson process is the stochastic process in which events occur continuously and independently of one another. In its spatial application, Poisson point processes have the property of no interaction between points or complete spatial randomness. Poisson processes can be spatially homogeneous or have a spatially inhomogeneous intensity correlated to the spatial variation of specific parameters. Homogeneous Poisson is used as reference processes when comparing against more structured point process models, in order to test the non-randomness of the spatial point process.

Interpolation Techniques

Interpolation refers to a set of methods aimed at modelling the value of a variable for which we only have a series of measurements or sampling points. Two main groups of interpolation techniques exist: deterministic and geostatistical. The former refers to techniques, such as Inverse Distance Weighting, Bilinear or Nearest Neighbour (see "Nearest Neighbour Techniques"), based on either the extent of similarity or the degree of smoothing of the samples. The outcomes of deterministic techniques is characterised by unpredictable estimation errors, due to uncertainty and variability of parameters. On the contrary, geostatistical predictions, such as those provided by kriging methods (see "Kriging"), are based on an underlying statistical model that can take additional information into account as explanatory variables and enable the prediction error to be estimated.

Kernel Algorithms for Density Estimation

Kernel Density Estimates (KDE) enable the modelling of the probability density function of a random variable, such as the population of certain finite data sample. KDE spreads the known quantity of the analysed phenomenon across the study region, according to a defined bandwidth parameter. This allows the creation of the density map by placing a "bump" (i.e. kernel) over each point in turn and adding all of them together in the final surface.

Kriging

Kriging is the most commonly used spatial prediction method for interpolation. While most methods use a weighted average of the sample values to generate the prediction, kriging determines these weights based on the semivariogram function. The general algorithm of kriging predictor assumes that the observations are generated by an isotropic, intrinsically stationary Gaussian process, so not all kind of data can be modelled by it. This is a problem that usually arises in disciplines such as archaeology, where we can observe an indiscriminate use of kriging, instead of the adequate tools, for modelling all kind of spatial phenomena.

Markov Random Field models

Markov random field models are 2D variation of markov chains. A Markov chain consists in values that are conditionally independent, i.e. each values depends exclusively on the previous one. In the case of spatial analysis, a Markov field is defined by a series of sites (points in the lattice) that are dependent on their 4 neighbours' values and which satisfy the Markov properties. The values of each point depend exclusively on their present state and not on the past or future chain of values of the point. This particular model is most useful to check for the palimpsest effect and study the temporal variations between samples.

Multiple-point geostatistics

The term multiple-point geostatistics refers to the use of training images to refine the predictions. The method is based on the assumption that the algorithm used for prediction can be trained by using other studies (images) of known distribution patterns. Thus, the spatial structure of points is not derived by a model but empirically derived from known case-studies.

Nearest Neighbour Techniques

This technique compares the observed distance between points and the expected distance for a homogeneous distribution. The resultant index explains the aggregation or inhibition pattern that the spatial distribution exhibits, taking into consideration only the distance between each element and its nearest neighbour.

An extension of this technique is the nearest neighbour interpolation, which uses the same principle to model the value of a variable for a whole region from a set of controls or sampling points. Rather than calculate an average value by some weighting criteria or generate an intermediate value, this method simply determines the nearest neighbouring pixel, and assumes the intensity (i.e. density) value for that location.

Pairwise Cost-based Distances Matrix

This is expressed as a table on which the cost-based distance between every point in a spatial distribution is represented. It allows the understanding of the spatial correlation between these entities when stationarity cannot be assumed as a premise. In other words, when the correlation between the elements of a spatial distribution cannot be assumed as a function of the Euclidean distance between them.

Point Pattern Analysis

The main goal of Point Pattern Analysis is the identification of the structural characteristics of a spatial point distribution. Inside this broad definition, we can find several statistical tools designed specifically for the study of spatial point patterns. The main goal of them is to derive insights from the location and context of a series of points that represents a real process under study. The most common approach is testing whether a covariate influences deviation from an assumed homogeneous or inhomogeneous model (see "Homogeneous/Inhomogeneous Poisson process") for the point pattern.

Principal Component Analysis

Principal Component Analysis (PCA) is a method used to compress the variation of a dataset, by identifying the main dimensions underlying its structure. PCA identifies correlated parameters that induce redundancy in the variation of the data, and it replaces these sets of parameters with "principal components" that summarize them. Principal Components (PC) are ranked according to their estimated influence on the overall variation of the dataset. The most common way of visualizing PCA results is by plotting the first- and second-ranked PCs: this enables PC1-PC2 covariation to be analysed and the residual variance in the dataset to be evaluated.

Pseudo-Euclidean Framework

When the correlation between the elements of a spatial point process cannot be assumed as a function of the Euclidean distance between them, its spatial relationship should be expressed through the use of alternative measurements. A recent approach to this problem proposes the modelling of a pseudo-Euclidean framework derived from the cost-based distances between the elements of the spatial distribution under study. That space presents the characteristics of a common Euclidean region, but the arrangement of the point pattern is transformed into a new one by non-metric multidimensional scaling of the original coordinates through its pairwise cost-based distances matrix.

Pseudo-genetic models

Pseudo-genetic models use a modified genetic algorithm to study predictions in an evolutionary framework. The predicted value thus becomes the result of a selection of best-fitted predictions obtained through multiple simulations of possible predictions. This is defined as an optimization technique that tend to minimise heterogeneity and is at present mostly used to model hydrologic systems.

Ripley's K-function

This is a widely used statistics to analyse interaction in point process data. It is commonly used as a test, where the null hypothesis is that the point process under consideration is a homogeneous Poisson process and the alternative is that the point process exhibits aggregation or segregation. These can refer to external processes affecting the point pattern (**first order effects**) or clustering/inhibitory behaviours (**second order effects**). Inhomogeneous Poisson process models are usually created for testing the existence of first order effects. The main advantage of Ripley's K is that it combines distance and frequency methods in order to identify not only dispersion rates but the measure of them in an increasingly range of scales. It can also be used to analyse the mutual interaction of different sets of marked points (bivariate K-function). Since the estimation of the degree of spatial interaction relies on visual assessment (interaction values are plotted against distance), the cumulative K-distribution is often transformed using the *L*-function to improve the readability of the plot.

Ripley-Rasson estimate

Ripley-Rasson estimator provides an estimation of the spatial window or region needed to compute certain statistics of point pattern analysis, such as the K-function. This window is obtained from data by computing a "convex hull" (i.e. the smallest region in which each line connecting any two points is completely within the region itself) and rescaling it according to the number of observations and vertices.

Simulated envelop

This is a Simulation-based method used in many applications of spatial statistics when sampling distributions are not valid or unavailable. Commonly, random permutations of a homogeneous process are compared with the data to assess hypothesis-based tests, for instance in Ripley's K-function.

Spatial autocorrelation

According to Tobler's first law of geography, the correlation between spatial values is inversely proportional to the distance between them. That is, correlation decreases when distance increases, or in other words, near events are more related than distant ones. This property of spatial features is known as spatial autocorrelation.

Standard Deviational Ellipse

A common way of measuring the trend for a set of points or areas is to calculate the standard distance separately in the x and y directions. These two measures define the axes of an ellipse encompassing the distribution of features. The ellipse is referred to as the standard deviational ellipse, since the method calculates the standard deviation of the x coordinates and y coordinates from the mean center to define the axes of the ellipse. The ellipse shows if the distribution of features is elongated and hence has a particular orientation.

Sum of squared-error of prediction (SSE)

In statistics, SSE (or Residual Sum of Squares - RSS) gives a measure of how empirical values deviate from the predicted values. The term residual indicates the difference between the observed value and the estimated value of the measurement of interest. Residuals are always referred to the statistical sample under study rather than to the general statistical population, thus constituting an observable estimate of the unobservable population error.

Variogram

Variogram, also know as semivariogram, represents the main tool for geostatistics, configuring a function that links the distance and the variance of a regionalized variable (the values of which present an autocorrelated spatial structure). The empirical estimation is commonly calculated by half the average of the squared difference between all the paired observations within the same discrete interval (usually called the lag or *h* vector). The length of such vector is arbitrary, and the estimation can be carried out omnidirectionally (for isotropic process) or in a specific direction (for anisotropic process). For variables that follow the intrinsic hypothesis (which variance is finite in the translation of *h*), the variogram is considered transitive, and present three interesting parameters: 1) **range**: the distance at which the variance do not increase anymore; 2) **sill**: the maximum semivariance (theoretically the same as the a priori variance of the variable); and 3) **nugget**: the semivariance shown in short distances (caused usually by sampling deficiencies or errors in the measurements). Some techniques, such as kriging interpolation methods, require a model of variogram in order to make predictions. For this purpose, a mathematical model is obtained through the empirical estimation (e.g. adjusted by the method of least squares).